CITIES OF RUSSIA

Introduction. The idea of my research is to help people, who want to change their place of residence. My family ran into with a problem - children are often sick, both sons are allergic. We realized that the problem is the environmental situation in the city where we live. Therefore, we decided to look for a city to change their place of residence. We do not want to leave the country. Only change the city. But there are many cities in Russia and it is very difficult to make the right choice. The study will help me analyze the cities and make the right choice. This study will be useful to those who want to move to another city; real estate agencies of individual cities, in addition, it may become a takeover for changes in troubled cities:

- People who plan to change their place of residence will be able to visually assess the capabilities of individual cities.
- Real estate agencies will be able to use the results of research to advertise real estate in individual cities.
- City leaders can get a clear picture of the state of their city.

I need a lot of different ***data***:

1. location of cities;
2. ecological situation in cities;
3. the population growth rate of the city (what is the trend?: more come or leave the city)
4. cost of housing.

As additional data, I need data on vacancies and the level of wages in cities. However, I decided that this information is for each user will be personal. People work in various industries and it is difficult to build recommendations for each industry in this study. Data sources:

- https://showdata.gks.ru/report/279006/
- https://www.domofond.ru/statya/ekologicheskiy_reyting_200_gorodov_rossii_za_2019_god/100219
- https://rosrealt.ru/cena/kvartira

The data obtained from various sources were combined into one dataframe (Pic.1). The data frame originally collected the following data:

- ✓ 2018 - 2012 - urban population migration for 2018-2012 (a positive value means that the population is arriving; a negative value means the population is decreasing)
- ✓ total - the total number of population migration for 6 years;
- ✓ mean - the average value of population migration over 6 years;
- ✓ populat - urban population in 2018;
- ✓ Mean_populate - share of population migration of the total population in 2018
- ✓ ecology 2019 and 2018 - the coefficient of the ecological situation of the city for 2019 and 2018 (the lower the value, the worse the environmental situation);
- ✓ cost room - the cost per square meter of housing in 2018;
- ✓ longitude, latitude - location coordinates

Note that the sample involved cities of Russia with a population of more than 100,000 people.

	city	2018	2017	2016	2015	2014	2013	2012	Total	Mean	populat	Mean_populat	ecology 2019	ecology 2018	cost room	longitude	latitude
0	Чебоксары	1580	242	4774	2845	2380	854	2042	14717	3679.25	492000	0.003211	7.2	7.1	41401.0	47.244960	56.130719
1	Абакан	1909	2256	2069	1820	1870	2336	1501	13761	3440.25	184000	0.010375	6.8	7.2	50749.0	91.438873	53.715290
2	Ижевск	495	979	167	-1457	2456	2370	996	6006	1501.50	648000	0.000764	6.8	6.7	44839.0	53.209417	56.866557
3	Казань	3171	3482	5527	3626	8791	8537	7359	40493	10123.25	1244000	0.002549	6.8	6.5	81421.0	49.124227	55.782355
4	Владикавказ	-2525	-2122	-1858	-2221	-1486	-2896	-3540	-16648	-4162.00	306000	-0.008252	7.0	6.5	35124.0	44.682110	43.024593
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
87	Красноярск	821	945	5848	6629	9620	12543	12890	49296	12324.00	1091000	0.000753	5.2	5.0	60449.0	92.872515	56.009097
88	Новороссийск	1481	2065	2268	3055	-1644	4257	3233	14715	3678.75	273000	0.005425	6.9	6.9	64483.0	37.769071	44.723958
89	Армавир	-399	-65	-432	-934	-490	7	782	-1531	-382.75	191000	-0.002089	7.6	7.9	43899.0	44.037701	40.156513
90	Краснодар	10573	11849	20208	17705	-4507	17060	16303	89191	22297.75	900000	0.011748	7.0	7.1	55483.0	38.976481	45.035257
91	Барнаул	64	-2760	-4094	-2752	-621	-221	5015	-5369	-1342.25	632000	0.000101	6.6	7.0	50035.0	83.778450	53.347402

92 rows × 17 columns

Pic.1.

The received data frame has been processed, the rows are empty. The properties of the data frame that are important for further processing are determined. The following properties were selected: the share of population migration for 2018; the cost per square meter of housing in 2018; environmental level in 2018 (Pic.2).

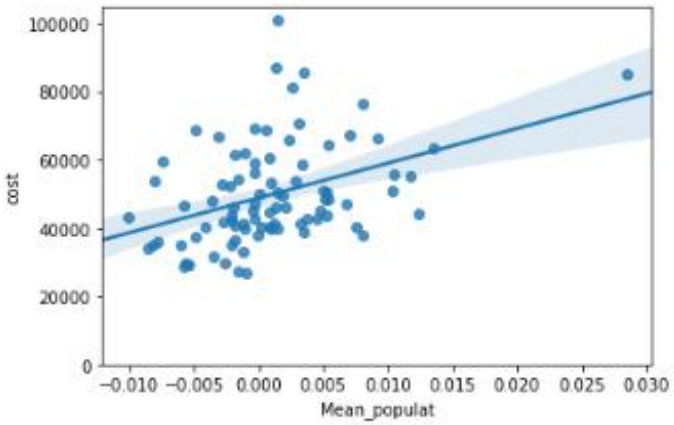
	city	Mean_populat	Ecology	cost	longitude	latitude
0	Чебоксары	0.003211	7.1	41401.0	47.244960	56.130719
1	Абакан	0.010375	7.2	50749.0	91.438873	53.715290
2	Ижевск	0.000764	6.7	44839.0	53.209417	56.866557
3	Казань	0.002549	6.5	81421.0	49.124227	55.782355
4	Владикавказ	-0.008252	6.5	35124.0	44.682110	43.024593
...	...	...	...	...	...	...
87	Красноярск	0.000753	5.0	60449.0	92.872515	56.009097
88	Новороссийск	0.005425	6.9	64483.0	37.769071	44.723958
89	Армавир	-0.002089	7.9	43899.0	44.037701	40.156513
90	Краснодар	0.011748	7.1	55483.0	38.976481	45.035257
91	Барнаул	0.000101	7.0	50035.0	83.778450	53.347402

87 rows × 6 columns

Pic.2.

Methodology

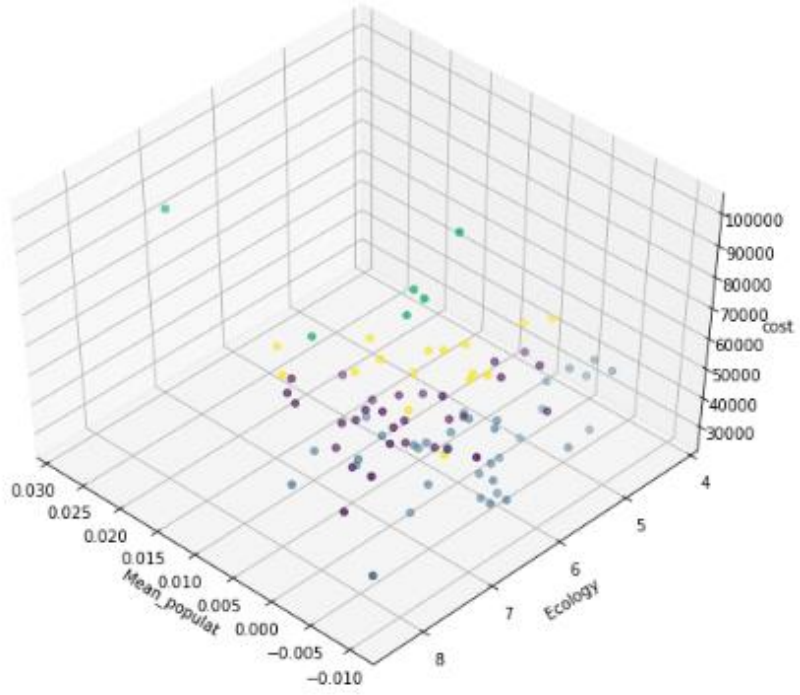
To analyze the datagram, the characteristics of an individual city were studied; when calculating the correlation, it turned out that there is a direct correlation between population migration and the cost of housing (Pic.3). The relationship between population migration and the level of environmental conditions is less weak, but it also exists.



Pic.3.

The Pearson Correlation Coefficient is 0.39808265294779915 with a P-value of P = 0.0001342924481422 9738

In this study, we are interested in which city is the best for moving, so it was decided to use segmentation of cities with the allocation of 4 segments. In this case, the K-Means Clustering method was used (Pic. 4).



Pic. 4.

We tried to divide the cities into segments with similar characteristics. There were few cities with medium or low cost of housing and good nature. But they exist. Therefore, further we need to determine the names of these cities and visualize information on the environmental situation and the cost of housing.

All cities were divided into 4 categories according to the ecological situation:

- ✓ Ecology <= 5.9 - Very large pollution
- ✓ 5.9 < Ecology <= 6.3 - High pollution
- ✓ 6.3< Ecology <= 6.9 - Elevated pollution

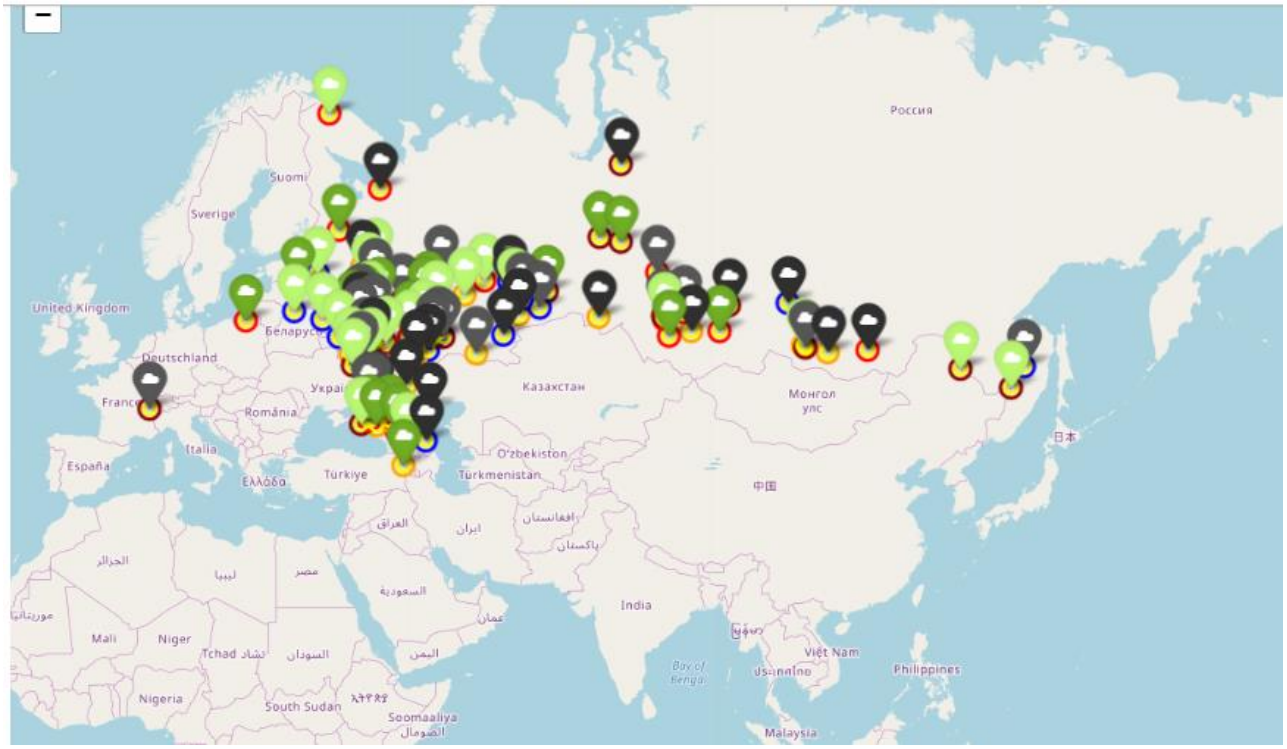
- ✓ Ecology >6.9 - Low pollution

We also divided all cities into 4 clusters according to the cost of housing:

- ✓ cost <= 40285 - "cluster 1"
- ✓ 40285 < cost <= 46625 - "cluster 2"
- ✓ 46625< cost <= 58772 "cluster 3"
- ✓ cost > 58772 "cluster 4"

View of ecology label	Means	View of cost label	Means
	Very large pollution		cluster 4
	High pollution		cluster 3
	Elevated pollution		cluster 2
	Low pollution		cluster 1

We set tag data for each city using folium (Pic. 5).



Pic. 5.

RESULTS

Analyzing the mark data on the map, we saw that for our family the following cities would be most preferable for moving:

- ✓ Vladimir,

- ✓ Saransk,
- ✓ Cheboksary,
- ✓ Pskov.

DISCUSSION

This study can be continued by adding additional characteristics of cities: crime rate; average salary, medical care. And initially, I planned so, but unfortunately I did not find data on these characteristics. But I will try to continue the work.

In conclusion, I want to note that I really liked to conduct this study, for a long time I was not fascinated by the study process. It was great, I really enjoyed it!